PANDA: POSE ALIGNED NETWORKS FOR DEEP ATTRIBUTE MODELING

Ning Zhang, Manohar Paluri, Marc'Aurelio Ranzato,
Trevor Darrell, Lubomir Bourdev

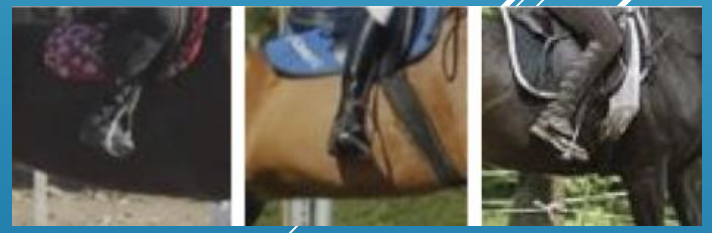
UC Berkeley, Facebook

Presented by Hossein Azizpour

WHAT ARE VISUAL ATTRIBUTES?

- My definition: an image quality which can be shared across ***proper subset*** of categories, subcategories, or instances ***of interest***
- Semantic
  - has-legs, is-furry, is-shiny,
- Non-semantic
  - Discriminative local mid-level visual features (almost analogous to shareable parts)
  - Discriminative global mid-level visual features (not done yet?!)
- Binary: has hat
- Relative: smiling face



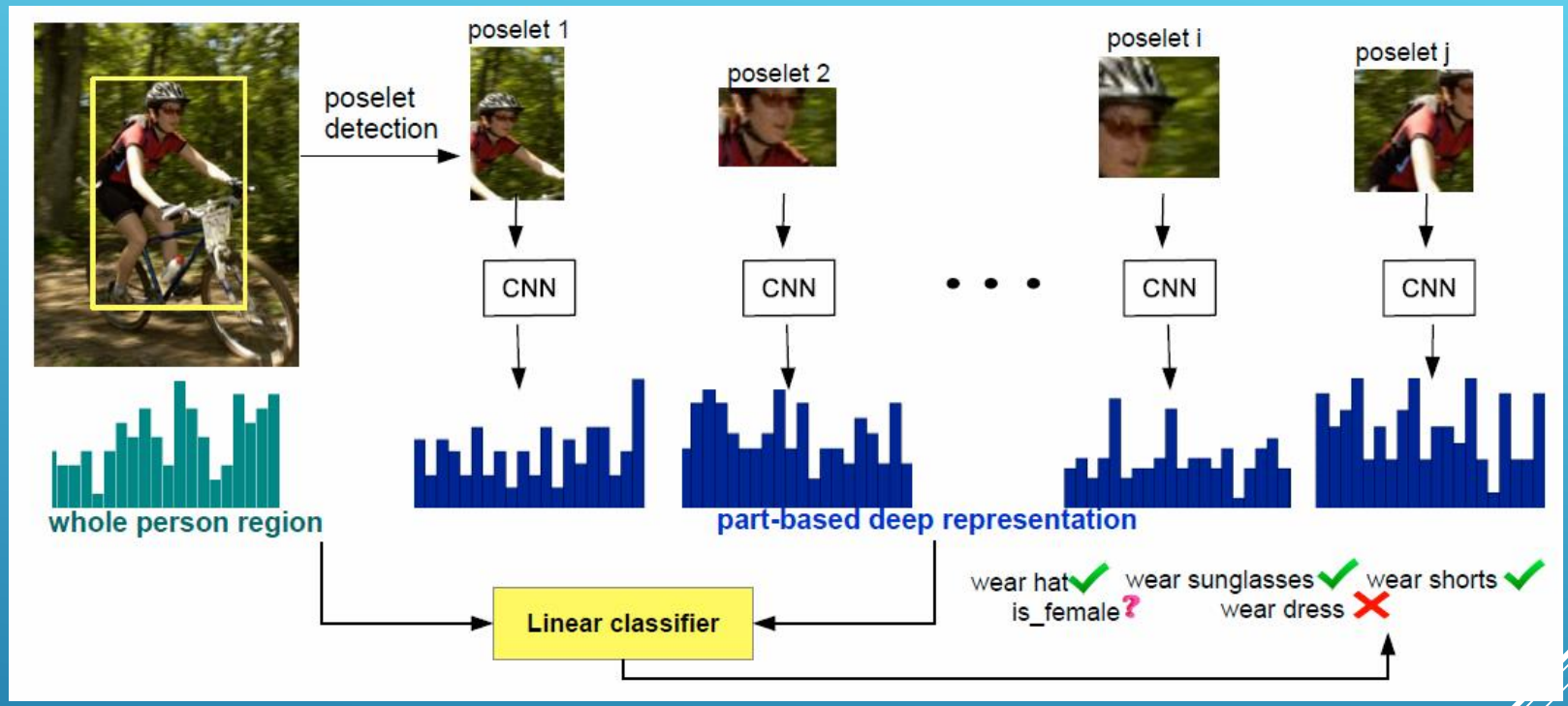




WHY BOTHER WITH ATTRIBUTES?

- Attributes enable
  - unknown object description
  - sample sharing (better generalization)
  - compressed representations

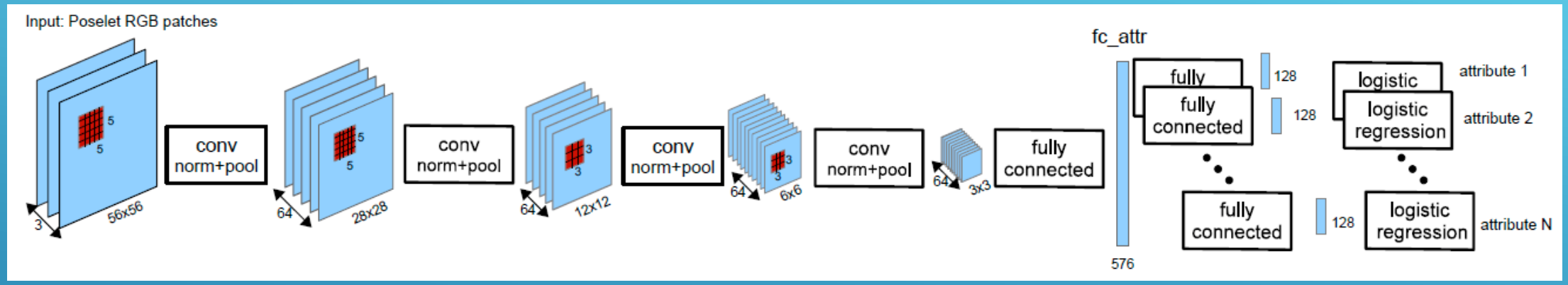
PANDA

- CNNs are good for learning discriminative features
- Poselets are good for making canonical representations



- CNN does not work well with low number of images
  - Pose normalized inputs
  - Related task with larger amount of images

POSELET CNN



- 64x64
- Jitter and flipping
- 25K additional dataset with additional attributes
- Branches out

RESULTS - FINAL

Attribute	male	long hair	glasses	hat	tshirt	longsleeves	shorts	jeans	long pants	Mean AP
Poselets[4]	82.4	72.5	55.6	60.1	**51.2**	74.2	45.5	54.7	90.3	65.18
DPD[27]	83.7	70.0	38.1	73.4	49.8	78.1	64.1	78.1	93.5	69.88
PANDA	**91.7**	**82.7**	**70.0**	**74.2**	49.8	**86.0**	**79.1**	**81.0**	**96.4**	**78.98**

Table 1: Attribute classification results on Berkeley Attributes of People Dataset as compared to the methods of Bourdev *et al.* [4] and Zhang *et al.* [27] .

- PANDA is good!
- DPD is Deformable Part Descriptors
- Poselets is the original paper

RESULTS – TRANSFER LEARNING

Attribute	male	long hair	hat	glasses	dress	sunglasses	short sleeves	baby	mean AP
Poselets150[4]	86.00	75.31	29.03	36.72	34.73	50.16	55.25	41.26	51.06
DPD[27]	85.84	72.40	27.55	23.94	48.55	34.36	54.75	41.38	48.60
DeCAF [8]	82.47	65.03	19.15	14.91	44.68	26.91	56.40	50.19	44.97
PANDA	**94.10**	**83.17**	**39.52**	**72.25**	**59.41**	**66.62**	**72.09**	**78.76**	**70.74**

Table 2: Average Precision on the Attributes25K-test dataset.

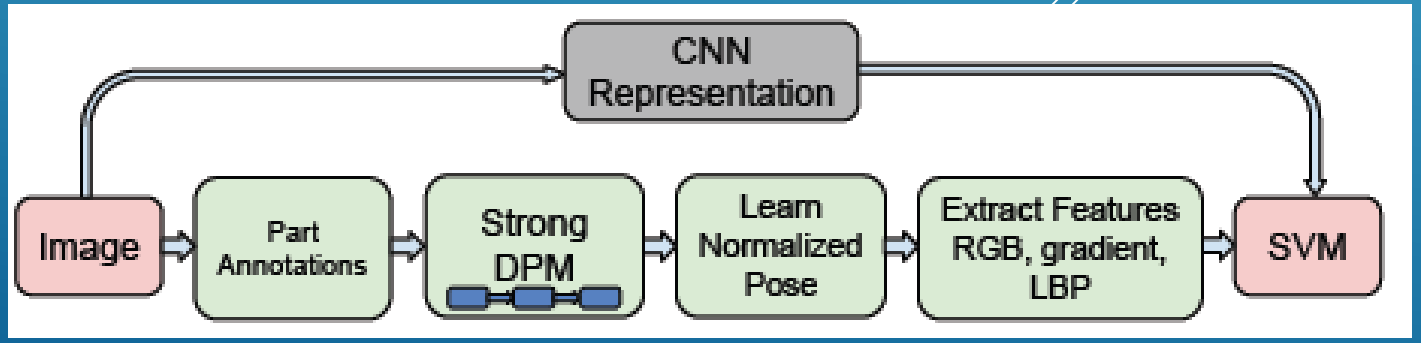
- DeCAF is off-the-shelf implementation

RESULTS – DIFFERENT METHODS

Attribute	male	long hair	glasses	hat	tshirt	longsleeves	short	jeans	long pants	Mean AP
DL-Pure	80.65	63.23	30.74	57.21	37.99	71.76	35.05	60.18	86.17	58.11
DeCAF	79.64	62.29	31.29	55.17	41.84	78.77	**80.66**	**81.46**	96.32	67.49
Poselets150 L2	81.70	67.07	44.24	54.01	42.16	71.70	36.71	42.56	87.41	58.62
DLPoselets	**92.10**	82.26	**76.25**	65.55	44.83	77.31	43.71	52.52	87.82	69.15
PANDA	91.66	**82.70**	69.95	**74.22**	**49.84**	**86.01**	79.08	80.99	**96.37**	**78.98**

Table 3: Relative performance of baselines and components of our system on the Berkeley Attributes of People test set.

- DL-Pure: all, top, middle, bottom 64x128
- Poselets150 L2: (HOG features, color histogram, skin tone and part masks)
- DL Poselets: RGB!
- PANDA: DL-PURE + DL Poselets
- Overfeat+SVM **70.78**



RESULTS - FACE ATTRIBUTES

- 13233 images
- 5749 people
- Cropped
- Centered
- Crowd sourcing: gender attribute only
- Same network is used as PANDA

Method	Gender AP
Simile [15]	95.52
FrontalFace poselet	96.43
PANDA	**99.54**

RESULTS - VIEWPOINTS

Partition	male	long hair	glasses	hat	tshirt	longsleeves	shorts	jeans	long pants	Mean AP
Frontal	92.55	88.40	77.09	74.40	51.69	86.84	78.00	79.63	95.70	80.47
Profile	91.42	59.38	37.06	69.47	49.02	84.61	85.57	82.71	98.10	73.04
Back-facing	88.65	63.77	72.61	72.19	55.20	84.32	74.01	86.12	96.68	77.06
All	91.66	82.70	69.95	74.22	49.84	86.01	79.08	80.99	96.37	78.98

Table 4: Performance of PANDA on front-facing, profile-facing and back-facing examples of the Berkeley Attributes of People test set.

- Frontal is better! (alignment is good!)

HIGH SCORING IMAGES

(a) Query: Women with long hair who wear glasses.

(b) Query: people who wear hats and glasses.

(c) Query: men with short pants and glasses.

Figure 6: Example of failure cases on Berkeley Attributes of People Dataset. On the left we show highest scoring failure cases for "wears t-shirt" and on the right – for "wears long sleeves".

FAILING PARTS